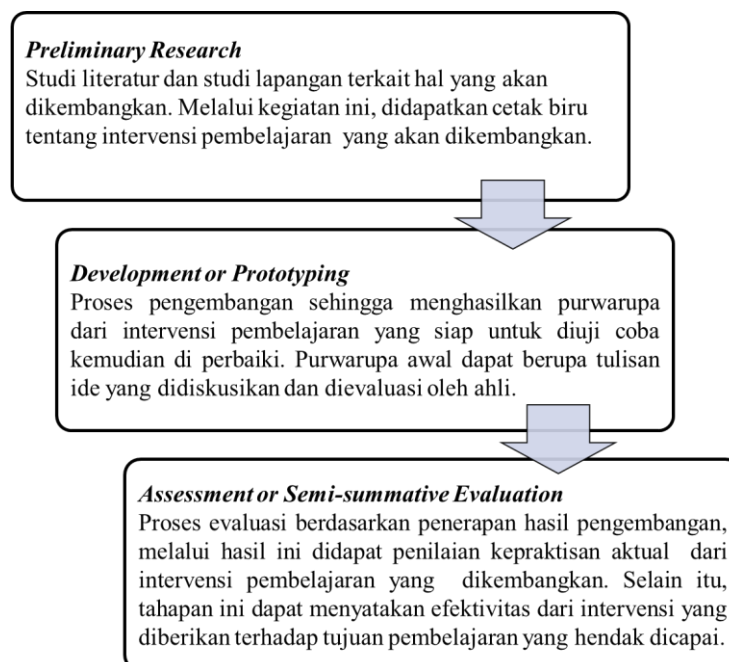


BAB III METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini bertujuan untuk mengembangkan *project task* dan rubrik penilaian keterampilan 4K. Desain penelitian yang digunakan adalah *educational design research* (EDR) yang merujuk pada Plomp (2013). Secara umum, terdapat dua tujuan dari EDR, yaitu sebagai penelitian pengembangan dan penelitian validasi. Penelitian ini menggunakan EDR dalam konteks penelitian pengembangan. Tujuan dari EDR dalam konteks pengembangan adalah mengembangkan solusi berbasis penelitian untuk masalah dalam praktik pendidikan. Definisi EDR dalam konteks ini adalah analisis sistematis serta desain dan evaluasi dari intervensi pembelajaran dengan dua tujuan yaitu: 1) menghasilkan solusi berbasis penelitian untuk masalah dalam praktik pendidikan, 2) meningkatkan pengetahuan tentang karakteristik dari intervensi tersebut beserta proses merancang dan mengembangkannya.

Terdapat tiga tahapan dalam EDR sebagai penelitian pengembangan. Ketiga tahapan beserta penjelasan singkat tentang aktivitas yang harus dilakukan pada setiap tahapan adalah sebagai berikut:



Gambar 3.1. Tahapan Pengembangan Plomp (2013)

1. Tahap Pertama: *Preliminary research*

Salah satu karakteristik dari EDR adalah melibatkan praktisi dalam kegiatan penelitian (Plomp, 2013). Pada tahap ini dikumpulkan informasi terkait pembelajaran dan penilaian keterampilan abad 21 terhadap guru fisika di SMA, SMK, dan MA menggunakan angket yang disebar dengan teknik daring. Setelah itu, untuk memperkuat informasi dari respon angket oleh guru, dilakukan analisis terhadap 60 artikel penelitian tentang keterampilan abad 21 pada mata pelajaran fisika untuk mendapatkan gambaran perkembangan penelitian keterampilan abad 21 di Indonesia. Kedua informasi ini akan menjawab pertanyaan penelitian pertama terkait cuplikan profil pembelajaran fisika dalam memfasilitasi keterampilan 4K di Indonesia

2. Tahap Kedua: *Prototyping*

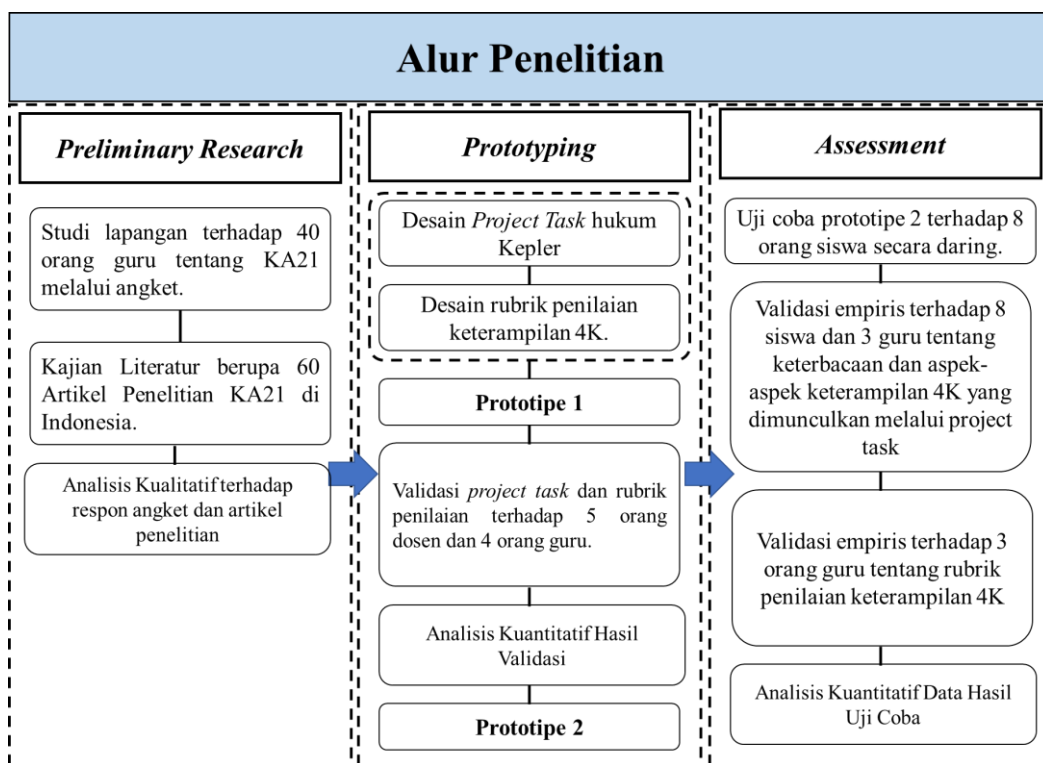
Teridentifikasi 4 keterampilan abad 21 yang diketahui oleh guru dan sering menjadi kajian penelitian di Indonesia, yaitu keterampilan berpikir kritis, kreativitas, kolaborasi, dan komunikasi. Empat keterampilan ini selanjutnya disebut dengan keterampilan 4K yang akan difasilitasi dengan pembelajaran berbasis proyek dengan kerangka acuan Boss (2015). Pemilihan pembelajaran berbasis proyek berdasarkan rekomendasi dari guru serta posisi penelitian melalui analisis artikel. Setelah teridentifikasi, dibuat desain *project task* serta rubrik penilaian keterampilan 4K, yang disebut dengan Prototipe I. Melalui tahap *prototype*, desain *project task* serta rubrik penilaian divalidasi oleh ahli dan praktisi. Hasil validasi pada tahapan ini berupa penilaian ahli dan praktisi terhadap item-item penilaian yang muncul pada aspek validasi terkait *project task* dan rubrik penilaian keterampilan 4K. *Project task* dan rubrik penilaian 4K yang telah divalidasi dan diperbaiki disebut dengan Prototipe II.

3. Tahap Ketiga: *Assessment*

Prototipe II dianalisis lebih lanjut menggunakan data empiris yang diperoleh dari siswa dan guru. *Project task* 1, 2, dan 3 diuji coba secara terbatas di salah satu SMA Negeri di Kabupaten Cianjur. Setelah siswa menyelesaikan ketiga *project task*, siswa diminta untuk mengisi angket untuk menilai keterbacaan dan aspek-aspek keterampilan 4K yang dimunculkan melalui *project task*.

Hasil ini ditujukan untuk memvalidasi *project task* berdasarkan penilaian siswa yang telah menyelesaikan *project task* tersebut. Selanjutnya, dilakukan validasi empiris terhadap rubrik penilaian keterampilan 4K. Transkripsi proses diskusi siswa diberikan kepada tiga orang guru kemudian dinilai menggunakan rubrik penilaian keterampilan 4K. Pola penilaian guru terhadap rubrik penilaian keterampilan 4K digunakan untuk memvalidasi rubrik itu sendiri.

Secara ringkas ketiga tahapan tersebut kemudian diilustrasikan dalam diagram alur penelitian sebagai berikut.



Gambar 3.2. Desain Penelitian

3.2. Partisipan dan Tempat Penelitian

Partisipan dalam penelitian ini terdiri dari dosen, guru, serta siswa. Pada tahap *preliminary research*, 40 orang guru fisika yang berasal dari SMA, SMK, MA, dan bimbingan belajar terlibat dalam pengisian angket terkait pembelajaran dan penilaian keterampilan 4K siswa pada pembelajaran fisika di sekolah. Sejumlah 40 orang guru tersebut adalah guru yang berkenan dalam mengisi angket yang disebar pada forum guru musyawarah guru mata pelajaran fisika (MGMP) di beberapa

kabupaten seperti Cirebon, Karawang, Tangerang, Banten, Bogor, Bandung, Cianjur, Tasikmalaya, Majalengka, Purwakarta, dan Ciamis. Diantara 40 orang guru yang mengisi angket tersebut juga pernah mengikuti *workshop* pembelajaran berorientasi keterampilan abad 21 yang diselenggarakan oleh prodi pendidikan fisika UPI pada tahun 2019. Melalui proses ini, diharapkan respon angket yang diberikan lebih representatif.

Pada tahap *prototyping*, partisipan yang dilibatkan untuk menilai desain *project task* dan rubrik penilaian keterampilan 4K adalah lima orang dosen dan empat orang guru. Lima orang dosen tersebut terdiri dari tiga orang dosen ahli pembelajaran berorientasi keterampilan abad 21, serta dua orang dosen astronomi. Empat orang guru yang dipilih berasal dari empat kabupaten yang berbeda yaitu Bogor, Bandung, Majalengka, dan Purwakarta. Pertimbangan peneliti dalam menentukan subjek guru sebagai penilai merujuk pada bidang penelitian yang pernah dilakukan masing-masing guru ketika menyelesaikan tugas akhir pada jenjang S1. Dua orang guru dengan pengalaman mengajar kurang dari 5 tahun berasal dari bidang penelitian tentang keterampilan abad 21. Sementara dua orang lainnya dengan pengalaman mengajar 5 sampai 10 tahun berasal dari bidang penelitian tentang penilaian pembelajaran.

Melalui tahap *assessment*, delapan orang siswa kelas X SMA Negeri 1 Warungkondang tahun ajaran 2019/ 2020 bertindak sebagai partisipan dalam uji coba terbatas *project task*. Delapan orang siswa tersebut merupakan bagian dari 89 orang siswa di tempat peneliti mengajar. Siswa tersebut dipilih berdasarkan pertimbangan peneliti atas aktivitas siswa selama pembelajaran daring berlangsung. Kedelapan siswa merupakan siswa yang paling aktif dan kooperatif selama pembelajaran daring pada masa pandemi COVID-19. Berdasarkan pertimbangan tersebut, peneliti berasumsi bahwa proses penyelesaian ujicoba *project task* dapat berlangsung dengan optimal. Selain itu, pada tahap ini dilibatkan tiga orang guru yang bertindak sebagai penilai proses diskusi siswa dalam menyelesaikan *project task*.

3.3. Teknik dan Instrumen Pengumpulan Data

Dalam rangka menjawab pertanyaan penelitian, diperlukan beberapa data yang mendasari jawaban atas pertanyaan tersebut. Data yang diperlukan untuk menjawab pertanyaan penelitian disajikan dalam Tabel 3.1

Tabel 3.1 Data Beserta Teknik Pengumpulan Data

Pertanyaan penelitian	Data yang diperlukan	Teknik Pengumpulan Data
Cuplikan profil pembelajaran fisika dalam memfasilitasi keterampilan 4K di Indonesia.	Respon guru tentang pembelajaran dan penilaian keterampilan abad 21.	Angket
	Perkembangan penelitian tentang keterampilan abad 21 pada mata pelajaran fisika kaitannya dengan pembelajaran berbasis proyek.	Dokumentasi Artikel Penelitian
Karakteristik desain <i>project task</i> dan rubrik penilaian keterampilan 4K berdasarkan ahli dan praktisi.	Validasi ahli dan praktisi terhadap desain <i>project task</i> dan rubrik penilaian keterampilan 4K.	Lembar Validasi
Kelayakan <i>project task</i> dan rubrik penilaian keterampilan 4K berdasarkan hasil uji coba.	Skor penilaian keterbacaan dan muatan keterampilan 4K dalam <i>project task</i>	Angket Penilaian
	Skor guru terhadap aspek-aspek keterampilan 4K siswa selama proses diskusi	Lembar Penilaian

3.3.1. Angket tentang Pembelajaran Fisika dan Kaitannya dengan Keterampilan Abad 21

Salah satu tujuan dari penelitian ini adalah mengetahui cuplikan profil pembelajaran fisika dalam memfasilitasi keterampilan 4K di Indonesia berdasarkan sudut pandang guru dan peneliti. Digunakan angket tentang pembelajaran dan

penilaian keterampilan abad 21 yang ditujukan terhadap guru, untuk mendapat informasi tentang pandangan guru tentang keterampilan abad 21, serta pengalaman guru dalam melatih dan menilai keterampilan abad 21 siswa dalam mata pelajaran fisika. Pertanyaan dalam angket dibuat dengan jenis *open-ended question*, diajukan dengan menggunakan angket dalam bentuk daring. Terdapat enam butir pertanyaan yang diajukan kepada guru, seperti ditunjukkan dalam Tabel 3.2.

Tabel 3.2. Butir Pertanyaan yang Diajukan terhadap Guru tentang Pembelajaran Fisika dan Kaitannya dengan Keterampilan Abad 21

No.	Butir Pertanyaan
1.	Apa yang bapak/ ibu ketahui tentang aspek-aspek keterampilan abad 21?
2.	Hal apakah yang pernah bapak/ ibu lakukan untuk melatih keterampilan abad 21 pada mata pelajaran fisika?
3.	Bagaimana cara yang pernah bapak/ ibu lakukan dalam menilai keterampilan abad 21 siswa pada mata pelajaran fisika?
4.	Menurut penilaian bapak/ ibu, bagaimana kondisi siswa/ siswi yang bapak/ ibu ajar berkaitan dengan keterampilan abad 21 mereka?
5.	Jika bapak/ ibu diminta untuk memberi rekomendasi, pembelajaran seperti apa yang dapat memfasilitasi keterampilan abad 21 siswa?
6.	Jika bapak/ ibu diminta untuk memberi rekomendasi, bagaimana teknik penilaian yang sesuai dalam menilai keterampilan abad 21 siswa di sekolah?

3.3.2. Angket Validasi Desain *Project Task* dan Rubrik Penilaian Keterampilan 4K

Project task dan rubrik penilaian keterampilan yang telah dibuat kemudian divalidasi oleh sembilan orang ahli dan praktisi. Angket validasi terdiri dari sepuluh butir pernyataan yang berfungsi menilai *project task* dan rubrik keterampilan 4K. Enam dari butir pernyataan yang merupakan item penilaian *project task* dibuat berdasarkan kriteria *task* yang baik berdasarkan Yang dan Liu (2016) dan Wulan (2018). Empat butir pertanyaan merupakan item penilaian rubrik keterampilan 4K berdasarkan kriteria rubrik yang baik menurut Wulan (2018). Contoh dari angket validasi ditampilkan pada Tabel 3.3.

Tabel 3.3. Contoh Item Penilaian dalam Lembar Validasi *Project Task* dan Rubrik Penilaian Keterampilan 4K

No.	Aspek Penilaian	Penilaian			
		4	3	2	1
1. <i>Project Task</i>					
1.a	<i>Project task</i> memfasilitasi siswa mempelajari konsep hukum Newton tentang gravitasi dan hukum Kepler.				
1.b	<i>Project task</i> memfasilitasi guru memperoleh informasi penting mengenai kemampuan siswa dalam hal keterampilan 4K.				
2. Rubrik Penilaian Keterampilan 4K					
2.a	Deskriptor keterampilan 4K siswa pada rubrik merupakan keterampilan penting untuk dinilai melalui observasi.				
2.b	Skala penskoran pada rubrik sudah menggambarkan tingkatan/ gradasi keterampilan 4K siswa serta memungkinkan untuk dapat diobservasi.				

Setiap item penilaian dinilai dengan empat skala penilaian, yaitu (1) Tidak Setuju, (2) Cenderung Tidak Setuju, (3) Cenderung Setuju, (4) Setuju. Skala penilaian ini berdasarkan rekomendasi dari Sumintono dan Widhiarso (2014) tentang hal-hal yang perlu dipertimbangkan dalam membuat opsi peringkat untuk responden. Skala penilaian tidak menggunakan opsi tengah, misalnya opsi *netral*, hal ini merujuk pada pernyataan Sumintono dan Widhiarso (2014) bahwa dalam fungsi pengukuran yang memerlukan penilaian responden, artinya ketegasan respons dianggap penting, maka opsi tengah sebaiknya tidak digunakan. Lembar validasi *project task* dan rubrik penilaian keterampilan 4K dapat dilihat secara lengkap pada Lampiran A.2.

3.3.3. Angket Penilaian Guru dan Siswa terkait Muatan Keterampilan 4K dalam Project Task

Pertanyaan panduan yang disajikan dalam *project task* dibuat berdasarkan deskriptor keterampilan 4K dengan kerangka acuan Boss (2015), melalui pertanyaan tersebut, diharapkan siswa dapat difasilitasi untuk melatih keterampilan 4K serta membantu siswa dalam menyelesaikan proyek yang diberikan. Delapan orang siswa yang mengerjakan seluruh *project task* diminta untuk mengisi angket penilaian siswa sebagai bahan refleksi kegiatan yang sudah siswa lakukan selama

mengerjakan *project task*. Selain itu, guru yang menilai kegiatan diskusi siswa juga menilai tentang muatan keterampilan 4K yang diharapkan muncul melalui pengerjaan *project task*. Setiap pernyataan (item penilaian dinilai) dibuat dalam empat skala penilaian, yaitu (1) Tidak Setuju, (2) Cenderung Tidak Setuju, (3) Cenderung Setuju, (4) Setuju. Sumintono dan Widhiarso (2014) menyatakan bahwa jenis pernyataan yang disajikan berkaitan dengan perilaku responden pada masa lalu/ telah dilewati tidak perlu menyediakan opsi tengah, misalnya *netral*. Dengan pertimbangan tersebut, dibuat empat skala penilaian tanpa opsi tengah. Contoh pertanyaan dalam angket penilaian muatan keterampilan 4K dalam *project task* disajikan dalam Tabel 3.4. Lebih lengkapnya, butir pernyataan penilaian muatan keterampilan 4K dalam *project task* terdapat pada Lampiran A.3.

Tabel 3.4. Contoh Butir Pernyataan dalam Angket Penilaian Guru dan Siswa terkait Muatan Keterampilan 4K dalam Project Task

No.	Pernyataan	Penilaian			
		4	3	2	1
1.	Melalui pengerjaan <i>project task</i> , saya dilatih untuk mencari dan menggunakan informasi dari berbagai sumber.				
2.	<i>Project task</i> memfasilitasi dalam menganalisis data, informasi, atau fakta untuk menyimpulkan sesuatu.				
3.	<i>Project task</i> melatih untuk menyampaikan pendapat, sudut pandang, atau solusi dari permasalahan.				

3.3.4. Angket Keterbacaan *Project Task*

Panduan atau pertanyaan yang muncul didalam *task* merupakan bentuk komunikasi guru terhadap siswa melalui komunikasi tulisan. Dengan demikian, panduan atau pertanyaan harus dipastikan dapat dipahami dengan baik oleh siswa dan guru. Dengan demikian, sebagai salah satu penunjang kelayakan dari *project task* yang dikembangkan, maka dilakukan pengujian *project task* dari aspek komunikasi, dalam penelitian ini disebut dengan angket keterbacaan. Pertanyaan dalam angket disusun berdasarkan keefektifan komunikasi secara tertulis melalui *handout* oleh Prabavathi dan Nagasubramani (2018). Contoh dari angket penilaian keterbacaan ditunjukkan pada Tabel 3.5

Tabel 3.5. Contoh Butir Pernyataan dalam Angket Keterbacaan

No.	Pernyataan	Penilaian			
		4	3	2	1
1. a	Narasi pada <i>project 1</i> dapat dipahami dengan mudah				
1. b	Jika terdapat hal yang tidak dipahami, pada bagian mana hal yang tidak dipahami?	Uraian:			
2. a	Tidak terdapat istilah/ kata yang tidak dimengerti di narasi <i>project 1</i>				
2. b	Jika terdapat istilah yang tidak dipahami, istilah apa yang tidak dimengerti?	Uraian:			

3.3.5. Lembar Penilaian Keterampilan 4K

Lembar penilaian keterampilan 4K digunakan oleh penilai untuk menuliskan hasil penilaian terhadap siswa berkaitan dengan keterampilan 4K selama penyelesaian *project task* berlangsung. Keterampilan yang dinilai adalah keterampilan berpikir kritis, kreativitas, kolaborasi, serta komunikasi. Deskriptor setiap keterampilan telah disediakan di dalam rubrik, dengan rentang skala penilaian 1-4. Hasil penilaian yang diberikan oleh tiga orang guru akan digunakan untuk memvalidasi rubrik penilaian keterampilan 4K. Contoh lembar penilaian ditampilkan pada Tabel 3.6.

Tabel 3.6. Contoh Lembar Penilaian Keterampilan 4K

Lembar Penilaian Keterampilan Berpikir Kritis dalam Pembelajaran Berbasis Proyek			
Kelompok yang dinilai :			
Anggota Kelompok :			
Kode Siswa	Skor yang diberikan	Deskriptor Berpikir Kritis yang Muncul	Kegiatan Diskusi yang Teramati
SI. 1			
SI. 2			
SI. 3			
SI. 4			

3.4. Teknik Analisis Data

Data yang terkumpul melalui instrumen penelitian kemudian dianalisis untuk menjawab pertanyaan penelitian dan membuat kesimpulan. Terdapat lima data yang dihasilkan melalui pengisian instrumen, yaitu: 1) respon guru terhadap angket pembelajaran dan penilaian keterampilan abad 21, 2) hasil validasi sembilan orang ahli, 3) hasil penilaian siswa dan guru terhadap muatan 4K yang muncul dalam *project task*, 4) hasil penilaian siswa dan guru terhadap keterbacaan *project task*, serta 5) penilaian guru terhadap rubrik penilaian keterampilan 4K. Berikut disajikan penjelasan tentang teknik analisis yang digunakan untuk masing-masing data hasil penelitian.

3.4.1. Analisis Angket tentang Pembelajaran Fisika dan Kaitannya dengan Keterampilan Abad 21

Thematic analysis menggunakan bantuan perangkat lunak NVivo Pro 12 dilakukan untuk menganalisis respon guru terhadap angket pembelajaran dan penilaian keterampilan abad 21 siswa. *Thematic analysis* adalah metode untuk menganalisis data untuk mengidentifikasi pola atau untuk menemukan tema melalui sintesis data yang telah dikumpulkan oleh peneliti (Heriyanto, 2018; Bag dan Calik, 2017; Nowell, dkk., 2017). Setiap respon yang diberikan oleh guru untuk setiap pertanyaan dikelompokkan (diberikan kode) berdasarkan kata kunci. Misalnya, untuk pertanyaan nomor 1 tentang pengetahuan guru mengenai keterampilan abad 21, maka dapat dikelompokkan berdasarkan aspek-aspek keterampilan abad 21 yaitu berpikir kritis, berpikir kreatif, penyelesaian masalah, dan penguasaan teknologi. Melalui pengelompokan ini, dapat diketahui pola jawaban guru sehingga dapat diambil kesimpulan.

3.4.2. Analisis Angket Validasi Desain *Project Task* dan Rubrik Penilaian Keterampilan 4K

Skor yang diberikan oleh validator untuk setiap item pada lembar validasi dianalisis menggunakan pemodelan *many-facets Rasch model* (MFRM) dengan

bantuan perangkat lunak Minifac versi 3.83.3. Interaksi antara *rater* (validator), *ratee* (angket validasi) serta item (item-item validasi) menjadi salah satu alasan penggunaan MFRM. Sebelum mengetahui hasil validasi terhadap *project task* dan rubrik penilaian keterampilan 4K, dilakukan analisis terhadap *rater severity/leniency*. *Rater severity/leniency* didefinisikan sebagai kecenderungan penilai untuk memberikan skor yang lebih rendah atau lebih tinggi dibandingkan dengan rata-rata skor dari penilai lain penilai lain, berpotensi merugikan hal yang sedang dinilai (Erguvan dan Dünya, 2020). Melalui analisis ini, didapat informasi tentang kualitas validator ketika memberikan penilaian terhadap *project task* dan rubrik penilaian keterampilan 4K. Untuk mengetahui *rater severity/leniency*, digunakan analisis *person fit*. *Person fit* adalah ringkasan statistik yang merepresentasikan persetujuan antara validator dengan respon yang diharapkan berdasarkan model Rasch (Engelhard dan Wind, 2018). Sumintono dan Widhiarso (2014) menyebutkan bahwa *person fit* dapat menunjukkan adanya kecenderungan pola tidak konsisten dari validator dalam memvalidasi.

Selain menganalisis *person fit*, salah satu hal yang penting untuk dianalisis adalah *item fit* (Boone, Staver, dan Yale, 2014). Jika *person fit* lebih fokus terhadap *rater* atau validator, *item fit* fokus pada item yang dinilai. Dalam hal ini, item adalah pernyataan yang ada dalam lembar validasi yang menyatakan poin-poin yang divalidasi oleh validator terhadap *project task* dan rubrik penilaian keterampilan 4K. Sumintono dan Widhiarso (2014) menyatakan bahwa analisis terhadap *item fit* dapat menunjukkan adanya pemberian skor yang keliru (misalnya karena eror atau menebak skor). Penggunaan *person fit* dan *item fit* dalam menganalisis hasil validasi sebuah instrumen diantaranya dilakukan oleh Wibisono (2016) dan Sasongko, dkk. (2020).

Tahapan analisis *person fit* dan *item fit* berdasarkan rekomendasi dari Boone, Staver, dan Yale (2014) dan Khine (2020) adalah sebagai berikut:

- a. Mulai dengan menginvestigasi skor *outfit MNSQ*. Skor *outfit MNSQ* lebih dipertimbangkan dibandingkan dengan *infit MNSQ* karena *outfit* lebih fokus dibandingkan dengan *infit* dalam hal mengkaji *outliers* (misalnya kesalahan pemberian skor karena menebak). Kriteria *outfit* dan *infit* yang baik secara umum memiliki skor 0,5 sampai 1,5 namun dalam kasus khusus *judgement*

atau tingkat persetujuan, digunakan kriteria 0,4 sampai 1,2 (Wright dan Linacre, 1994; Boone, Staver, dan Yale, 2014).

- b. Jika dengan *outfit MNSQ* terjadi *misfits* (nilai tidak memenuhi kriteria) pada *person/ item*, maka dilakukan dengan analisis nilai *Outfit ZStd*, kemudian tentukan keputusan yang diberikan berdasarkan interpretasi skor *ZStd*. Kriteria *ZStd* berdasarkan Boone, Staver, dan Yale (2014) serta Sumintono dan Widhiarso (2015) adalah sebagai berikut.

Tabel 3.7 Interpretasi Nilai *Standardized Value (ZStd)*

Standardized Value (ZStd)	Implikasi untuk Pengukuran
≥ 3	Data tidak diharapkan jika sesuai dengan model (secara sempurna). Namun dengan sampel yang besar, ketidaksesuaiannya mungkin lebih kecil.
2,0 – 2,9	Data tampak tidak dapat diprediksi
-1,9 – 1,9	Data mempunyai perkiraan yang logis (dapat diprediksi)
≤ -2	Data terlalu mudah diprediksi

Untuk mengetahui hasil penilaian yang diberikan validator terhadap *project task* dan rubrik keterampilan 4K, digunakan nilai *observed average* dan *fair average*. Selain itu, digunakan juga ukuran *rater-agreement*. Hal ini mengacu pada Linacre (dalam Boone, Staver, dan Yale, 2014) bahwa untuk *rater-agreement* yang bertujuan meninjau persetujuan *rater* terhadap item yang dinilai, dapat menggunakan *rater-agreement percentage*. Interpretasi *rater-agreement percentage* ditunjukkan oleh Tabel 3.8 sebagai berikut.

Tabel 3.8 Interpretasi *Rater-Agreement Percentage*

<i>Rasch Agreement</i>	Interpretasi
<i>Observed < Expected</i>	Menunjukkan ketidaksepakatan, biasanya terjadi dengan penilai yang tidak terlatih.
<i>Observed \approx Expected</i>	Penilai bertindak secara independen.
<i>Observed > Expected</i>	Normal untuk penilai terlatih.

<i>Observed</i> >> <i>Expected</i>	Penilai tidak memberi peringkat secara independen. Mungkin ada tekanan untuk setuju dengan penilai lain.
<i>Observed</i> > 90%	Penilai berperilaku seperti mesin pemeringkat.

3.4.3. Angket Penilaian Guru dan Siswa terkait Muatan Keterampilan 4K dalam *Project Task*

Seperti dijelaskan sebelumnya bahwa angket penilaian guru dan siswa terkait muatan 4K dalam *project task* digunakan untuk memperkuat validasi *project task*. Hasil validasi sebelumnya, seperti dijelaskan pada poin 3.4.2 menjelaskan penilaian *project task* dari sudut pandang ahli dan praktisi, sementara hasil validasi ini menjelaskan penilaian guru dan siswa yang terhadap *project task*. Guru dan siswa dapat menilai *project task* berdasarkan pengalaman dalam menyelesaikan *project task* yang didalamnya memuat keterampilan 4K. Dalam validasi ini, yang bertindak sebagai *rater* adalah guru dan siswa, *ratee* adalah angket muatan keterampilan 4K dalam *project task*, sementara item yang dinilai adalah deskriptor keterampilan 4K yang disajikan dalam angket.

Pola analisis yang sama seperti poin 3.4.2 dilakukan terhadap respon guru dan siswa terhadap angket penilaian muatan 4K. Digunakan pemodelan MFRM dengan bantuan perangkat lunak Minifac 3.83.3. Pertama, dilakukan analisis terhadap penilai, yaitu guru dan siswa melalui *person fit* dengan meninjau nilai *Outfit MNSQ*. Jika nilai *Outfit MNSQ* tidak masuk ke dalam rentang yang ditentukan, digunakan nilai *ZStd*. Setelah seluruh penilai dinyatakan konsisten dalam memberikan penilaian (sesuai dengan model), maka dilakukan analisis terhadap *item*. Hasil penilaian guru dan siswa terhadap muatan keterampilan 4K dalam *project task* diketahui dari *observed average* dan *fair average*.

3.4.4. Angket Keterbacaan *Project Task*

Angket keterbacaan *project task* bertujuan untuk memvalidasi *project task* berdasarkan efektivitas komunikasi terhadap siswa. Efektivitas komunikasi dalam hal ini berarti bahwa informasi atau panduan dalam *project task* dapat dipahami oleh pengguna yaitu siswa. Ukuran ini digunakan untuk mengantisipasi *project task* tidak dipahami oleh siswa ketika digunakan. Guru yang menilai proses diskusi

siswa juga dilibatkan dalam menilai keterbacaan *project task*. Dalam hal ini, yang bertindak sebagai *rater* adalah guru dan siswa, *ratee* adalah angket keterbacaan *project task*, sedangkan item adalah poin-poin keterbacaan yang dinilai oleh guru dan siswa. Digunakan pemodelan MFRM dengan bantuan perangkat lunak Minifac versi 3.83.3 untuk menganalisis hasil angket keterbacaan *project task*. Proses serta hal yang dianalisis dari respon guru dan siswa terhadap angket keterbacaan sama seperti proses menganalisis angket validasi serta angket muatan 4K.

3.4.5. Lembar Penilaian Keterampilan 4K

Lembar penilaian keterampilan 4K bertujuan untuk menilai validitas dari rubrik penilaian keterampilan 4K. Analisis validitas skala peringkat adalah pengujian yang dilakukan untuk memverifikasi apakah peringkat (rating) pilihan yang digunakan membingungkan bagi responden atau tidak (Sumintono dan Widhiarso, 2014). Rubrik penilaian keterampilan 4K adalah instrumen penilaian keterampilan 4K siswa yang terobservasi selama diskusi dalam menyelesaikan *project task*, diharapkan setiap penilai memberikan nilai yang sama untuk deskriptor keterampilan 4K yang muncul dari satu orang siswa. Artinya, penilaian dapat menghasilkan kesimpulan tentang tingkat keterampilan 4K siswa dalam kelompok yang dinilai. Meskipun demikian, Sumintono dan Widhiarso (2014) menyebutkan bahwa pada pengukuran melalui observasi, sumber variasi skor bertambah dengan masuknya penilai (*rater*) dalam pengukuran.

Pengujian skala peringkat dalam rubrik penilaian keterampilan 4K dilakukan dengan menganalisis nilai rata-rata observasi. Jika nilai rata-rata observasi meningkat dari skala terkecil yaitu skor 1 ke skala terbesar yaitu skor 4, maka peningkatan nilai logit menunjukkan hasil yang konsisten (Sumintono dan Widhiarso, 2014; Wibisono, 2016). Hal ini menunjukkan bahwa rubrik penilaian keterampilan 4K sudah bisa membedakan siswa dengan level keterampilan 4K yang tinggi maupun rendah. Sementara untuk nilai *Andrich threshold*, yang merupakan sebaran nilai yang diberikan responden, jika nilai ini bergerak dari angka yang negatif menuju angka positif secara berurutan, maka empat skala penilaian yang diberikan sudah valid bagi penilai.

Ukuran lain untuk menentukan kualitas dari rubrik penilaian keterampilan 4K yaitu validitas dan reliabilitas. Reliabilitas dari instrumen menjelaskan seberapa jauh pengukuran yang dilakukan berkali-kali akan menghasilkan hasil yang sama (Sumintono dan Widhiarso, 2014). Disisi lain, kecenderungan untuk memperoleh hasil pengukuran yang berbeda-beda akan selalu muncul, rubrik yang reliabel tidak selalu menghasilkan skor yang sama persis pada penilaian yang berulang, akan tetapi perbedaan skor tersebut haruslah kecil dan masih dalam batas toleransi. Parameter yang digunakan dalam MFRM untuk menentukan reliabilitas rubrik adalah *item reliability*, Sumintono dan Widhiarso (2015) menginterpretasikan nilai *item reliability* sebagai berikut.

Tabel 3.9 Interpretasi Nilai Reliabilitas untuk Instrumen

Nilai	Interpretasi
< 0,67	Lemah
0,67 – 0,80	Cukup
0,80 – 0,90	Bagus
0,91 – 0,94	Bagus Sekali
> 0,94	Istimewa

Validitas instrumen adalah seberapa jauh pengukuran oleh instrumen dapat mengukur atribut yang seharusnya diukur (Sumintono dan Widhiarso, 2014). Dalam konteks pengukuran menggunakan rubrik, atribut yang diukur adalah keterampilan 4K. Salah satu parameter dalam MFRM yang dapat digunakan untuk menentukan validitas dari instrumen adalah unidimensionalitas. Interpretasi dari persentase unidimensionalitas oleh Sumintono dan Widhiarso (2014) ditunjukkan oleh Tabel 3.10 sebagai berikut.

Tabel 3.10 Interpretasi Persentase Unidimensionalitas

Persentase	Interpretasi
>20%	Terpenuhi
>40%	Bagus
>60%	Istimewa